STIC-ILL

No

3783432

From: Steadman, David (AU1652)
Sent: Wednesday, November 28, 2001 4:43 PM
To: STIC-ILL
Subject: Literature/ILL Request

Art Unit: 1652
Office: 10D-04
Mailbox: 10C-01 M3
Case Serial #:09/459,573

Please provide the following references:

1) J Mol Microbiol Biotechnol 1999 Nov;1(2):327-36
The LysE superfamily: topology of the lysine exporter LysE of Corynebacterium glutamicum, a paradyme for a novel superfamily of transmembrane solute translocators.
Vrljic M, Garg J, Bellmann A, Wachi S, Freudl R, Malecki MJ, Sahm H, Kozina VJ, Eggeling L, Sajier MH Jr, Eggeling L, Saier MH Jr.

2) Mol Microbiol 1996 Dec;22(5):815-26
A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum.
Vrljic M, Sahm H, Eggeling L.

Thank you,
David Steadman

footnote states that yahn family has been renamed to RhtB

J. Mol. Microbiol. Biotechnol. (1999) 1(2): 327-336.

JMMB Research Article

The LysE Superfamily: Topology of the Lysine Exporter LysE of *Corynebacterium glutamicum*, a Paradyme for a Novel Superfamily of Transmembrane Solute Translocators

M. Vrljic[1], J. Garg[2], A. Bellmann[1], S. Wachi[2], R. Freudl[1], M. J. Malecki[2], H. Sahm[1], V. J. Kozina[2], L. Eggeling[1]*, and M. H. Saier Jr.[2]

[1]Institut für Biotechnologie,
Forschungszentrum Jülich GmbH,
52428 Jülich, Germany
[2]Department of Biology,
University of California at San Diego,
La Jolla, CA 92093-0116, USA

Abstract

In *Corynebacterium glutamicum* the LysE carrier protein exhibits the unique function of exporting L-lysine. We here analyze the membrane topology of LysE, a protein of 236 amino acyl residues, using PhoA- and LacZ-fusions. The amino-terminal end of LysE is located in the cytoplasm whereas the carboxy-terminal end is found in the periplasm. Although 6 hydrophobic domains were identified based on hydropathy analyses, only five transmembrane spanning helices appear to be present. The additional hydrophobic segment may dip into the membrane or be surface localized. We show that LysE is a member of a family of proteins found, for example, in *Escherichia coli*, *Bacillus subtilis*, *Mycobacterium tuberculosis* and *Helicobacter pylori*. This family, which we have designated the LysE family, is distantly related to two additional protein families which we have designated the YahN and CadD families. These three families, the members of which exhibit similar sizes, hydropathy profiles, and sequence motifs comprise the LysE superfamily. Functionally characterized members of the LysE superfamily export L-lysine, cadmium and possibly quarternary amines. We suggest that LysE superfamily members will prove to catalyze export of a variety of biologically important solutes.

Introduction

An unusual bacterial transport protein was recently characterized biochemically, physiologically, and molecularly (Broer and Krämer, 1991; Vrljic *et al.*, 1995; 1996). This carrier is LysE from the Gram-positive bacterium *Corynebacterium glutamicum*. It exports L-lysine, thereby regulating the intracellular concentration of this amino acid. The need for this function is surprising as the L-lysine biosynthetic pathway in *C. glutamicum*, as in other bacteria, is strictly regulated (Nakayama, 1985; Eggeling, 1994). However, L-lysine export is essential in environments containing peptides: in the presence of low concentrations of lysine-containing peptides and upon deletion of the export carrier gene *lysE*, exceptionally high cytoplasmic concentrations (>1M) of L-lysine build up leading to bacteriostasis. The specific export function provided by LysE is also a prerequisite for the production of L-lysine with *C. glutamicum*, mutants of which are used for the production of about 3.5×10^5 tons of this amino acid per year (Leuchtenberger, 1996). LysE therefore provides a new function in regulating a cytoplasmic amino acid concentration, and it is the first example of export as a target for the improvement of industrial amino acid production.

In addition to the unique physiological function of LysE, its structure is unusual. The carrier consists of 236 amino acyl residues (Vrljic *et al.*, 1996), and it exhibits six hydrophobic domains that could correspond to six transmembrane helical spanners typical of many polytopic membrane transport proteins. LysE lacks significant sequence similarity to known export translocators which are constituent members of the identified 12 families catalyzing efflux of organic molecules and cations (Saier, 1994; Saier *et al.*, 1994; Paulsen *et al.*, 1996). It therefore represents a novel family of proteins distinct from all other established families of transporters (Vrljic *et al.*, 1996). The current veritable flood of genome sequencing activities has resulted in the realization that numerous putative LysE homologues are encoded on bacterial chromosomes and plasmids. LysE evidently represents the tip of an iceberg of a novel class of transporters which may exhibit unique physiological functions analogous to that of LysE.

In the present paper we report investigations into the membrane topology of LysE using the well established methodology of alkaline phosphatase and β-galactosidase fusion constructions. Additionally, we searched the DNA and protein databases for the presence of LysE homologues. A thorough analysis revealed that LysE is a member of a novel superfamily of carriers that consists of three discrete families. The functionally characterized members of this superfamily exclusively catalyze solute export.

Results

The LysE Family

Currently sequenced members of the LysE family are presented in Table 1. The LysE protein of *C. glutamicum* is the only functionally characterized member of this family.

Received January 29, 1999; revised April 27, 1999; accepted May 4, 1999.
*For correspondence. Email l.eggeling@fz-juelich.de;
Tel. +49 2461 61 5132; Fax. +49 2461 61 2710.

Table 1. Protein Members of the LysE Family

Abbreviation	Name or description in database	Organism	Size (no. residues)	Database and accession no.
LysE Cgl	Lysine exporter protein	*Corynebacterium glutamicum*	236	gbX96471
Orf1 Mtu*	Hypothetical 20.9 KD protein CY20G9.14	*Mycobacterium tuberculosis*	212	spQ11154
Orf2 Mtu	Hypothetical 20.8 KD protein CY39.33C	*Mycobacterium tuberculosis*	199	spQ10871
YggA Eco	Hypothetical 23.2 KD protein in SBM-FBA intergenic region	*Escherichia coli*	211	spP11667
YggA Asa*	Lysine export homologue	*Aeromonas salmonicida*	206	gbU65741
YggA Ahy*	YggA homologue gene product	*Aeromonas hydrophila*	206	gbX89469
Orf Bsu	Hypothetical protein YisU	*Bacillus subtilis*	220	gbY09476
Orf Hpy	Conserved hypothetical integral membrane protein	*Helicobacter pylori*	210	gbAE000585

* Based on the DNA sequences; the amino acid sequences of these three proteins (used in the reported analyses) were modified from those reported in the data base entries as follows:
Orf1 Mtu: Eleven residues were added to the amino terminus of the sequence reported in the database. The initiation codon codes for valine.
YggA Asa: Fourteen residues were added to the amino terminus of the sequence reported in the database. The initiation codon codes for methionine.
YggA Ahy: A probable frameshift sequencing error in the 3' region of the gene was identified approximately at position 2241. The correct reading frame was deduced by comparison with the aligned homologues.

```
                                          *          * *        *                    * *
Orf Hpy     (1)                    MFVVFIEGFGLAISLCAAVGAQSLFIVERGMARNYV    FLICALCFMCDIVLM
YggA Eco    (1)                    MFSYYFQGLALGAAMILPLGPQNAFVMNQGIRRQYH    IMIALLCAISDLVLI
Orf2 Mtu    (1)                    MNSPLVVGFLACFTLIAAIGAQNAFVLRQGIQREHV    LPVVALCTVSDIVLI
YggA Asa    (1)                    MFATTLQGFTLGLAMIIPIGAQNAFVLSRGIHRNHH    LLAATLCCLCDLILI
Ygga Ahy    (1)                    MFATTLQGFTLGLAMIIPIGAQNAFVLSRGIHRNHH    LLTATLCCLCDLVLI
Orf1 Mtu    (1)                    VPLQVFVGFVAMMTLKVAIGPQNAFVLRQGIRREYV    LVIVALCGIADGALI
LysE Cgl    (1)                 MVIMEIFITGLLLGASLLLSIGPQNVLVIKQGIKREGL    IAVLLVCLISDVFLF
Orf Bsu     (1)  MTISLMIEKKCERSFFSMNAIIHGIVLAFGLILPLGVQNVFIFQQGALQKHIWRALPAVISASVCDTLLI

Consensus                          M------GF-L---LI--IG-QNAFV--QGI-R---   L----LC---D--LI


                 *                    *
Orf Hpy     (52)  SMGVFGVGAYFAKNLYLSLSLNLFGALFTGFYAFLALKTLFQTF                      KKKQV
YggA Eco    (52)  CAGIFGGSALLMQSPWLLALVTWGGVAFLLWYGFGAFKTAMSSN                      IELAS
Orf2 Mtu    (52)  AAGIAGFGALIGAHPRALNVVKFGGAAFLIGYGLLAARRAWRP                      VALIP
YggA Asa    (52)  GIGVFGGANLLAASPIGLALLTWGGVLFLCWFGIRSLRSAWQGQ                     GAALA
YggA Ahy    (52)  GIGVFGGANLLAASPIGLALLTWGGVLFLGWFGIRSLRSAWRGQ                     GAKLA
Orf1 Mtu    (52)  AAGVGGFAALIHAHPNMTLVARPGGAAFLIGYALLAARNAWRP                      SGLVP
LysE Cgl    (54)  IAGTLGVDLLSNAAPIVLDIMRWGGIAYLLWFAVMAAKDAMTNKVEAPQIIEETEPTVPDDTPLGGSAVA
Orf Bsu     (71)  VLAVAGVSVIVQELPVFETVMMAGGFLFLLYMG    WVTWNIR                      PNTSQ

Consensus         --GV-G---L--A-P--L-----GG--FL---G--A---AW---     '                 -----


                                                  ***   *      *          *          *
Orf Hpy     (101)  QTPKKLSLKKT            LLFTLGVTLLNPQVYLEMVFLIGASALSFNLAQKFVFLAGTLSAALSW
YggA Eco    (101)  AEVMKQGRWKI            IATMLAVTWLNPHVYLDTFVVLGSLGGQLDVEPKRWFALGTISASFLW
Orf2 Mtu    (100)  SGATPVRLAEV            LVTCAAFTFLNPHVYLDTVVLLGALANE HSDQRWLFGLGAVTASAVW
YggA Asa    (101)  DSPRLMGVKSV            LAMTLGVTLLNPHVYLDTLMLLGSFGSQFAEPLRPAFAAGAMLASLVW
YggA Ahy    (101)  DSPQLMGVKSV            LAMTLGVTLLNPHVYLDTLMLLGSPGSQPAEELRSAFAAVAMLASLVW
Orf1 Mtu    (100)  SESGPAALIGV            VQMCLVVTFLNPHVYLDTVVLIGALANE ESDLRWFFGAGAWAASVVW
LysE Cgl    (124)  TDTRNRVRVEVSVDKQRVWVKPMLMAIVLTWLNPNAYLDAFVFIGGVGAQYGDTGRWIFAAGAFAASLIW
Orf Bsu     (116)  NEKHTFTPKKQ            AAFAAAVSLLNPHAILDTIGVIGTSSLQYSGLEKWLFMAACIAVSWIW

Consensus          ----------V            -----L-VT-LNPHVYLDT--L-G----Q-----R--F-AGA--AS--W


                                                  *
Orf Hpy     (160)  LLLLCTLSLRYGSKLLNNQKIFMGVNLFVTAIMGTLSVTLFRDFLALLSKT
YggA Eco    (160)  FFGLALLAAWLAPR LRTAKAQRIINLVVGCVMWFIALQLARDGIAHAQALFS
Orf2 Mtu    (158)  FATLGFGAGRL RGLFTNPGSWRILDGLIAVMMVALGISLTVT
YggA Asa    (160)  FYSLAFGAAALSPW LARGRVQQAIDTIVGLIMLGLALQLASG  ALLAS
YggA Ahy    (160)  FYSLAFGAVVLSPW LARSRQQQAIDTIVGFTMLGLALQLASG  ALLAS
Orf1 Mtu    (158)  PAVLGFSAGRLQPF FATPAAWRILDALVAVTMIGVAVVVLVTSPSVPTANVALII
LysE Cgl    (194)  FPLVGFGAAALSRP LSSPKVWRWINVVVAVVMTALAIKLM      LMG
Orf Bsu     (175)  FISLAI AGRLFQTIDTSGRLMLIVNKCSAAVMWAAAGYFGV    SLFCN

Consensus          F--L-F-A--L---  L------------V---M--LA--L---   -L---
```

Figure 1. Complete multiple alignment of the sequences of the proteins that comprise the LysE family. The TREE program of Feng and Doolittle (1990) was used to derive the multiple alignments shown here and in subsequent figures. Asterisks above the alignment indicate fully conserved residues while residues conserved in a majority of the proteins are presented in the consensus sequence below the alignment. Residue numbers in each protein are provided in parentheses following the protein abbreviation at the beginning of each line. Abbreviations of the proteins are as indicated in Table 1.

The eight proteins found are of about the same size (199-236 residues) with the *C. glutamicum* protein being the largest. Probable sequencing errors and incorrect choices of initiation codons were detected in the genes encoding proteins from the two *Aeromonas* species as well as *Mycobacterium tuberculosis* (Orf1), and these led to incorrect protein sequence entries in the databases. These sequences were altered in accordance with the sequence of the characterized LysE protein and predictions based on the complete multiple alignment. The sequences used are shown in the alignment presented in Figure 1.

This alignment served to generate the average hydrophobicity plot shown in Figure 2A. Six peaks of hydrophobicity were observed with the potential to form membrane spanning α-helices. All of these regions occur without gaps in the multiple alignment with the exception of the divergent sequence of *B. subtilis* which exhibits a four-residue gap near the end of hydrophobic region #3. As transmembrane hydrophobic regions are usually better conserved than loop regions of integral membrane proteins, the results are consistent with the suggestion that most of these regions are transmembrane.

The largest hydrophilic loop of LysE is present between predicted hydrophobic regions #3 and #4. This loop is unique to LysE of *C. glutamicum*, since it is larger than that of its homologues and since it exhibits a high percentage of V and R residues (45%) as follows: A V A T D T R N R V R V E V S V D K Q R V. The Vs and Rs are underlined, and alternating hydrophilic residues are presented in bold face. Similar highly charged repeat sequences are found in loop regions of other transporters (Paulsen and Saier, 1997; Eng *et al.*, 1998).

Members of the LysE family show striking similarity throughout their lengths (Figure 1). As indicated by the asterisks, 16 residues are fully conserved. All but one of the six fully conserved glycyl residues are located in the first half of the alignment. The most conserved regions are present in hydrophobic regions #1, #2 and #4. These regions served for the derivation of signature sequences for the LysE family proteins, as follows:

1) (SAP) (LIV) G (PAV) Q (NS) (LIVA) (FL) (LIV) (LIVMF) X (QR) G

2) (CA) X (LIVM) (SAC) D (LIVTG) (LIVFA) L (LIVMF) X_2 (GA) X_2 G

3) (CATM) (LIVA) (GAV) (LIVF) (TS) (WFL) L N P (HNQ) (AV) (YI) L (DE)

(X = any residue; alternative residues at any position are indicated in parentheses).

The phylogenetic tree for the LysE family proteins is shown in Figure 3A. As expected for closely related orthologues, the two *Aeromonas* proteins cluster tightly together. The two *M. tuberculosis* paralogues are also similar in sequence, suggesting that these two proteins arose by a late gene duplication event that occured after Gram-positive bacteria diverged from Gram-negative bacteria. It is noteworthy that phylogenetic distances separating the proteins do not correlate with the

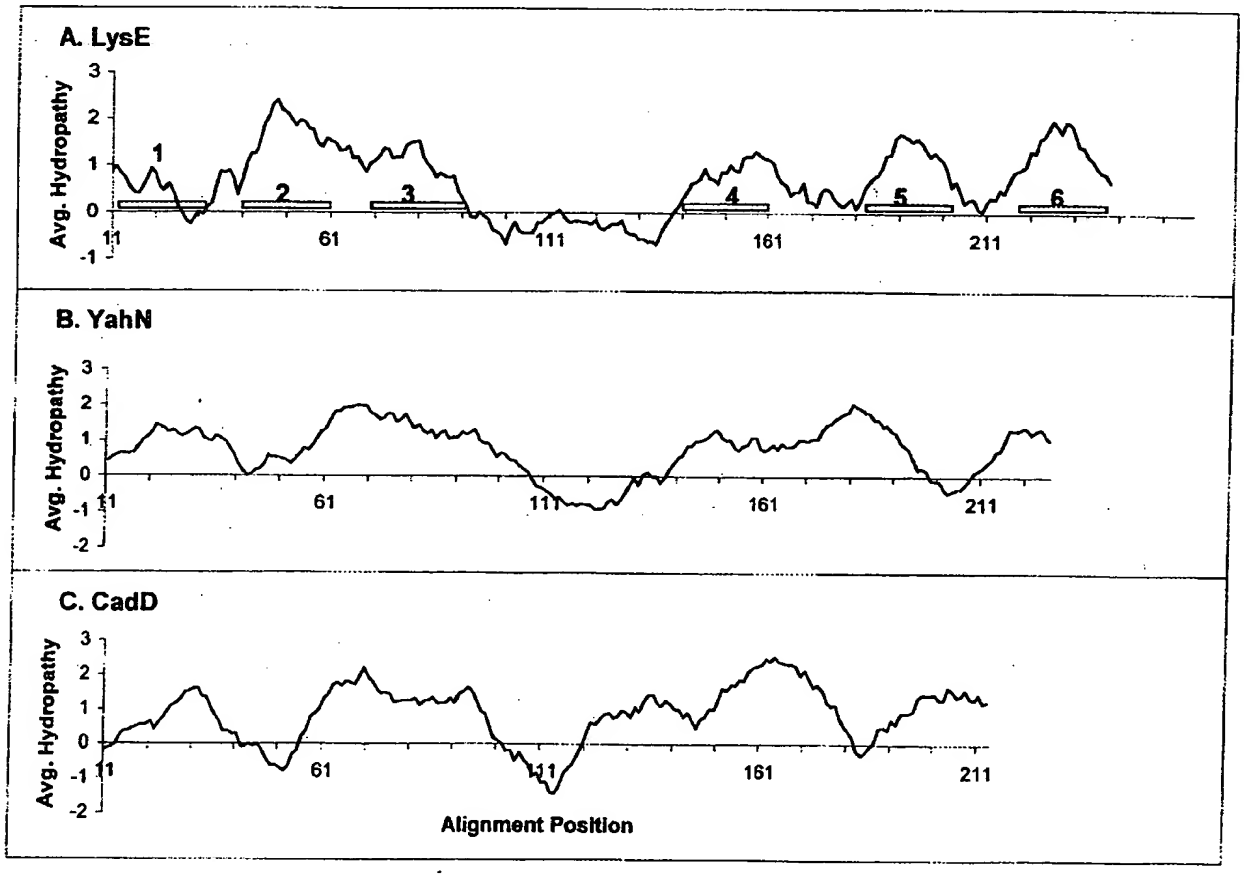


Figure 2. Average hydrophobicity plots for (A) the LysE family, (B) the YahN family, and (C) the CadD family. The average hydrophobicity plots were based on hydropathy values of individual amino acids as described by Kyte and Doolittle (1982).

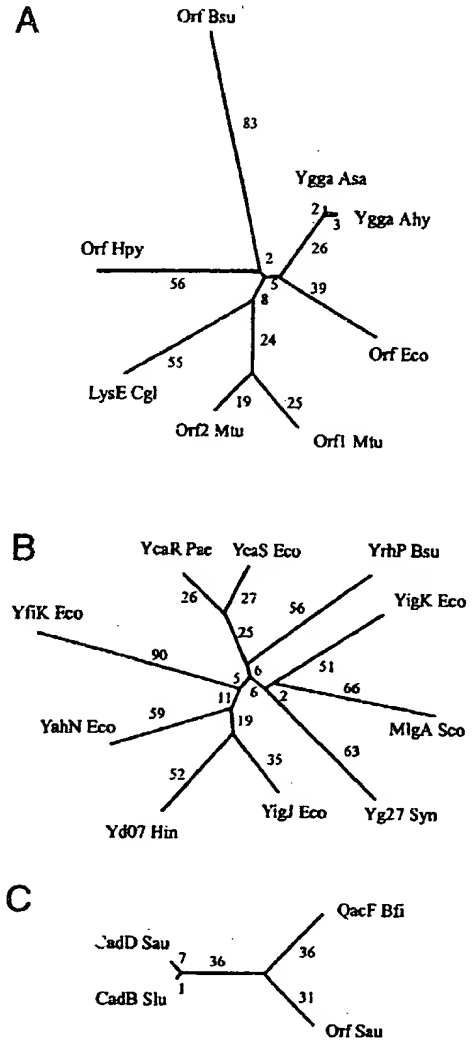


Figure 3. Phylogenetic trees for (A) the LysE family, (B) the YahN family, and (C) the CadD family. The trees were constructed using the TREE program (Feng and Doolittle, 1990; Saier, 1994). Branch lengths are approximately proportional to phylogenetic distance. Abbreviations for the proteins are as indicated in Tables 1-3.

phylogenies of the organisms. This fact suggests that the LysE family includes paralogues of dissimilar substrate specificities and functions.

Analysis of LysE-Fusion Proteins

The predicted positions of the 6 hydrophobic regions within LysE of *C. glutamicum*, based on the hydropathy analysis of the LysE family (Figure 2A) are shown in Figure 4 and numbered #1 to #6. Their putative start and termination points are also specified. Since *lysE* can be functionally expressed in *E. coli* (not shown), the well established alkaline phosphatase gene fusion technique was used to assay for the disposition of the polypeptide chain in the membrane (Manoil and Beckwith, 1986). Twenty two fusions were made to carboxy terminally truncated LysE molecules. As expected, assay of the respective recombinant *E. coli* CC118 strains carrying the *lysE"phoA* gene fusions gave either a blue or a white phenotype on LB plates containing the chromogenic alkaline phosphatase substrate 5-bromo-4-chloro-indolyl phosphate. The data for the quantitative assay of alkaline phosphatase specific activity are given in Figure 4. Fusions at positions 99, 111, 112, 113, 114, 115, 120, 130, 143, 154, 164, 183, and 211 (numbers correspond to the last amino acid of LysE in the LysE"PhoA polypeptide), resulted in nearly no alkaline phosphatase activity, comparable to the control containing the leaderless PhoA polypeptide. High activities resulted with all fusions made in the amino terminal parts of the protein as well as to locations 171 and 236 of LysE. In addition to the *phoA* fusions, *lacZ* fusions were also made. However, sequencing revealed that the correct constructs were only present in 2 of the 5 fusions produced. These 2 fusions at position 111 and 211 of LysE resulted in a blue phenotype in *E. coli* CC118 with 5-bromo-4-chloro-3-indolyl-ß-D-galactoside as substrate. They thus exhibit ß-galactosidase activity at fusion sites where the corresponding PhoA fusions did not cause any activity.

In order to establish that the LysE"PhoA fusion proteins had been synthesized and were present, Western blot analyses were conducted (Figure 5). The observed sizes of fusion proteins 111, 112, 114, and 115 are similar to the predicted sizes (61.9 kDa - 62.5 kDa). However, with the other fusions, visible quantities of the full-length fusion proteins were not detected, and the intensities of the PhoA-antibody positive proteins did not correlate with the PhoA specific activities. For instance, although fusions 59, 67, and 75 resulted in high specific PhoA activities (Figure 4), they did not result in visible proteins of the expected size of 55.5, 56.5, and 57.6 kDa, respectively. This suggests a low steady state level of the membrane-inserted fusion protein, possibly due to its susceptibility to rapid degradation. The instability of topology probe fusions has been documented previously (Sarsero and Pittard, 1995). Due to the degradation products present in the extracts analysed and the increased size observed for these proteins with increasing fusion-length, we suggest that these fusion proteins were synthesized and that the low activities observed were not due to poor expression.

LysE Topological Model

The high alkaline phosphatase activity observed with PhoA fused to the last amino acyl residue 236 suggested that the carboxyl end of the LysE polypeptide is localized to the periplasm. This fact determines the orientation of hydrophobic region #6 (Figure 6). In accordance with this proposed orientation is the fact that alkaline phosphatase is not translocated into the periplasm when fused at position 211, whereas ß-galactosidase fused at this position is active. The predicted loop region between hydrophobic regions #5 and #6 (loop 5:6) is therefore directed towards the cytoplasm. The PhoA fusion at position 171 is blue, localizing loop 4:5 to the periplasmic side. The white PhoA fusions at positions 99-130 in loop 3:4 and the blue LacZ fusion at position 111 place this unusually large loop of the LysE polypeptide in the cytoplasm. We therefore assign the amino terminal end of hydrophobic region #3 to the periplasmic side of the membrane. In accordance with this orientation is the translocation of PhoA when fused at positions 67 and 75. This assignment in the carboxy terminal part of LysE fits with a predicted carrier model exhibiting alternating loops and transmembrane spanning helices where each of the four hydrophobic regions #3 to #6 spans the membrane once (Figure 6).

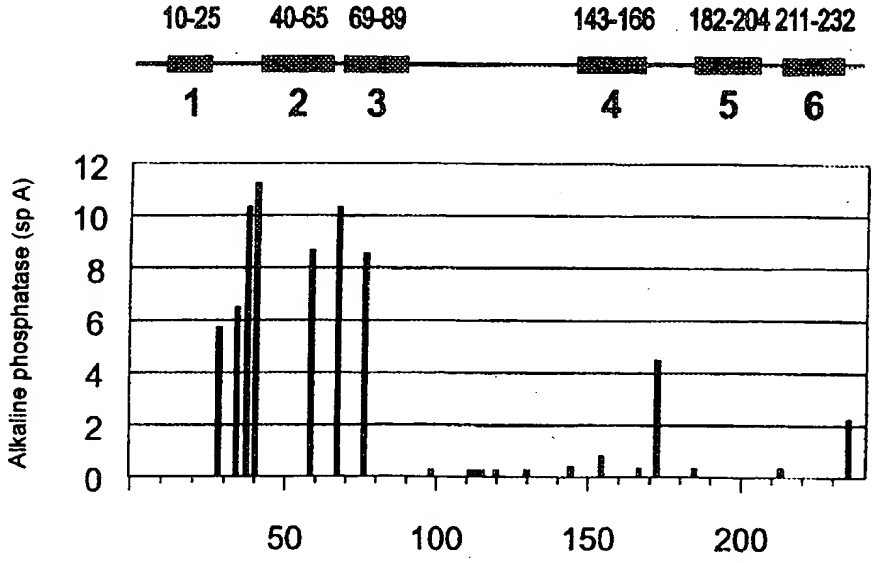
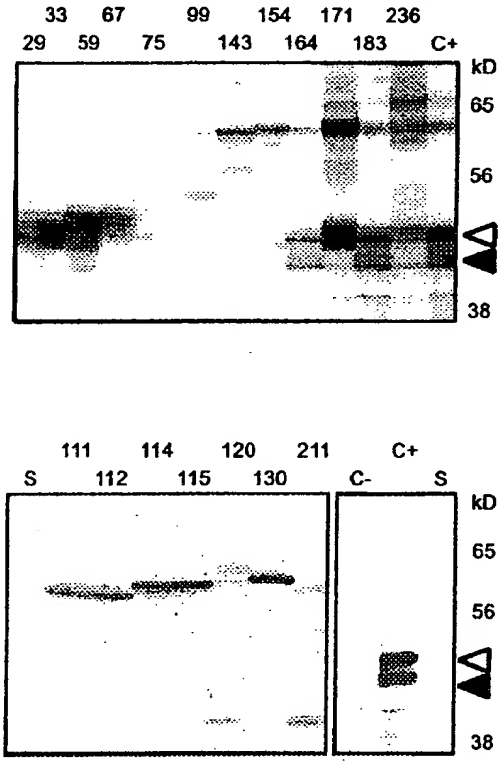


Figure 4. Hydrophobic regions of the *C. glutamicum* LysE protein and alkaline phosphatase activities of cells expressing LysE'PhoA fusion proteins. At the top of the figure, the predicted helicies of LysE are given together with their start and end points. The bars in the lower part of the figure give the specific alkaline phosphatase activities in µmol/min/mg (protein). They are located at the fusion sites of LysE with the sequence numbering given on the x-axis.



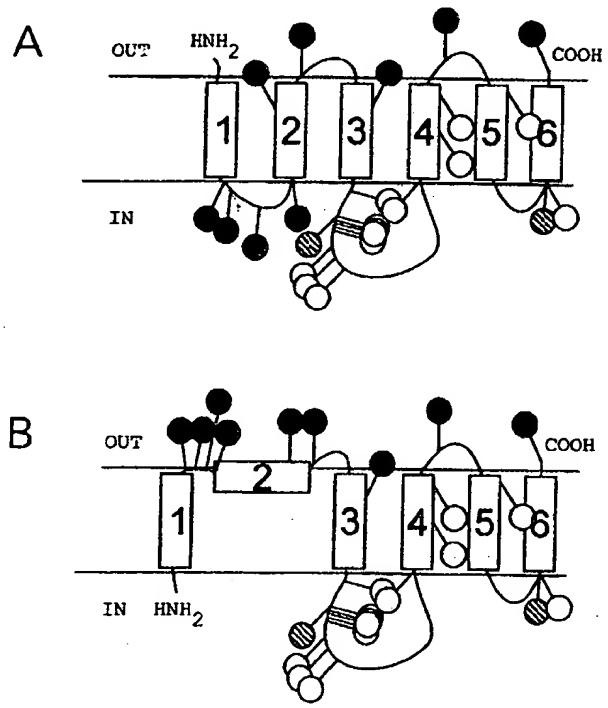


Figure 5. Western blot analysis of cells expressing *lysE'phoA* fusions. Cell extracts were separated on a 10% SDS polyacrylamide gel, and proteins were probed using antibody directed against PhoA. The individual fusion products are given with numbers indicating the last LysE-specific amino acid in the fusion protein. C+, extracts of *E. coli* CC118 pPA4 encoding pre-PhoA. C-, extract from the plasmidless strain. Pre-PhoA protein is marked by an open arrow head, and mature PhoA by a closed arrow head. S represents the protein standards with molecular weights in kilodaltons, given on the right.

Figure 6. Models for the two-dimensional topology of LysE of *C. glutamicum*. Models were predicted from (A) the average hydropathy plot, and (B) the PhoA/LacZ fusion analyses. The flags are located according to the assigned helices as given in Figure 4 and according to the fusion point. The dark and white flags indicate PhoA fusions resulting in high and low alkaline phosphatase activity, respectively. The two LacZ fusions obtained exhibiting high β-galactoside activity are given as hatched flags.

Table 2. Protein Members of the YahN Family

Abbreviation	Name or description in database	Organism	Size (no. residues)	Database and accession no.
YahN Eco	Hypothetical 24.8 KD protein in *betT-prpR* intergenic region	*Escherichia coli*	223	spP75693
YigJ Eco	Hypothetical 22.5 KD protein in *recQ-pldB* intergenic region	*Escherichia coli*	206	spP27846
Yd07 Hin	Hypothetical protein HI1307	*Haemophilus influenzae*	210	spQ57320
YcaR Pae	Hypothetical 23.3 KD protein in *carA-carB* intergenic region	*Pseudomonas aeruginosa*	216	spP38102
YeaS Eco	Hypothetical 23.2 KD protein in *gapA-rnd* intergenic region	*Escherichia coli*	212	spP76249
YrhP Bsu	Hypothetical protein in *aapA-sigV* intergenic region	*Bacillus subtilis*	210	gbZ99117
Yg27 Syn	Hypothetical 22.0 KD protein Slr1627	*Synechocystis PCC6803*	206	spP74343
MlgA Sco*	MlgA protein	*Shewanella colwelliana*	206	pirS23222
YigK Eco*	Hypothetical 15.4 KD protein in *recQ-pldB* intergenic region (F138)	*Escherichia coli*	204	spP27847
YfiK Eco	Hypothetical 21.2 KD protein in *srmB-ung* intergenic region	*Escherichia coli*	195	spP38101

*The sequences of these two proteins have been modified from those reported in the databases, based on the DNA sequences and the complete multiple alignment for the YahN family, as follows. MlgA: 24 residues were added to the amino terminus, and 32 residues were added to the C-terminus. These changes reflect (1) an incorrect selection of initiation codon, and (2) a sequencing error in the 575 nucleotide sequence region as reported under acc# gbX67020, respectively. YigK: 68 residues were added to the amino terminus. These changes reflect (1) an incorrect selection of initiation codon, and (2) a sequencing error in the 4092 nucleotide sequence region as reported under acc# gbU00096 (AE000458), respectively.

```
                                                          *                                *    *
YigJ Eco   (1)              MLMLFLTVAMVH   IVALMSPGPDFFFVSQTAVSRSRKEAMMGVLGITCGVMVWAGI
YigK Eco   (1)             MTLEWNFAYLLTS   IILSLSPGSGAINTMTTSLNHGYRGAVASIAGLQTGLAI  HI
YcaR Pae   (1)          MPTLGITDFWTYVLGV   VFVILLPGPNSLFVLATSAQRGVATGYRAACGVFLGDAVLMLL
Yg27 Syn   (1)          MLTMGWSNILSLFGAM   LIILAALPSLSVLTVSSKSASGGFIHGLFAALGVVLGDIIFILI
YahN Eco   (1) MMQLVHLFMDEITMDPLHAVYLTVGLFVITFFNPGANLFVVVQTSLASGRRAGVLTGLGVALGDAFYSGL
YeaS Eco   (1)         MFAEYGVLNYWTYLVGA   IFIVLVPGPNTLFVLKNSVSSGMKGGYLAACGVFIGDAVLMFL
Yd07 Hin   (1)                MMLNLIIVH   LFGLMTPGPDFFYVSRMAASNSRRNTVCGILGITLGIAFWGML
MlgA Sco   (1)             MSEHILIAVFLPT  PFFVSITPGMCMTLAMTLGMSIGVRRTLWMMVGELAGVALVAIA
YrhP Bsu   (1)               MHSLLAYIPIA   AMMVIIPGADTMLVMKNTLRYGPKAGRYNILGLATGLSFWTVI
YfiK Eco   (1)             MTPTLLSAFWTYT   LITAMTPGPNNILALSSATSHGFRQSTRVLAGMSLGFLIVMLL

Consensus                  -------------   ------PG-----V-----S-G---------G---G-A-----

                                         *
YigJ Eco  (56) ALLGLHLIIEKMAWLHTLIMVGGGLYLCWMGYQMLRGA L    KKEAVSAPAPQVELAKSGRSFLKGLL
YigK Eco  (55) VLVGLGTLFSRSVIAFEVLKWAGAAYLIWLGIQQWRAA G    AIDLKSLASTQ     SRRHLFQRAVF
YcaR Pae  (60) SALGVASLLKAEPMLFIGLKYLGAAYLFYLGVGMLRGAWR    KLRNPEATAGQAEQVDVHQRFRQALL
Yg27 Syn  (60) ALWGLAFLRGAMGDFFVILKYISGIYLSWLGINTIRA       KVNNQSLAKVDVK   SLSSSFSAGLL
YahN Eco  (71) GLFGLATLITQCEEIFSLIRIVGGAYLLWFAWCSMRRQ S    TPQMSTLQQPISA   PWYVFFRRGLI
YeaS Eco  (61) AWAGVATLIKTTPILFNIVRYLGAFYLLYLGSKILYAT L    KGKNSEAKSDEP    QYGAIFKRALI
Yd07 Hin  (53) SMLGLAVLFVTIPALHGVIMLLGGSYLAYLGFLMARSKKY    AKFESHSDTEFNQQTTIKKEILKGLL
MlgA Sco  (58) AVMGVASMMLNYPQLFDILKWVGGLYLGYIGISMWRA       KGKMANLDNTSSQ ISNRALITQGFV
YrhP Bsu  (55) AILGLSVVIAKSVILFTTIKYLGAAYLIYLGVKSFFAKSMFSLDDMQSQAKNMASSPKRYYKTSFMQGSL
YfiK Eco  (57) CAGISFSLAVIDPAAVHLLSWAGAAYIVWLAWKIATS       PTKEDGLQAK          PISFWASFA

Consensus      ---GLA-L------LF------G--YL--LG----R-- -    -------------  ------F--GL-

YigJ Eco (121) TNLANPKAIIYFGSVFSLFVGDNVG TTARWGI FAL IIVETLAWFTVVASLFALPQMR RGYQRLAKW
YigK Eco (116) VNLTNPKSIVFLAALFPQFIMPQQP QLMQYIV LGVTTIVVDIIVMIGYATLAQRIALW IKGPKQMKA
YcaR Pae (126) LSLSNPKAILFFISFFIQFVDPGYAYPGLSFLV LAVILELVSALYLSFLIFTGVRLAAWFRRRQRLAAG
Yg27 Syn (121) ITLADQKAVLFYLGFLPTFVDVNNI AYLDIAV IILTAILTVGGVKIFYAFLAHRSGLLISRQNK  RI
YahN Eco (134) TDLSNPQTVLFFISIFSVTLNAETP TWARLMAWAGI  VLASIIWRVFLSQAFSLPAVR RAYGRMQRV
YeaS Eco (123) LSLTNPKAILFYVSFFVQFIDVNAPHTGISFFI LAATLELVSFCYLSFLIISGAFVTQYIRTKKKLAKV
Yd07 Hin (119) VNLSNAKVVVYFSSVMSLVLVNIT  EMWQIILAFAV IVVETFCYFYVISLIFSRNIAK RLYSQYSRY
MlgA Sco (120) TAIANPKGWAFMISLLPPFISVDQA IAPQLMVLLSIIMMTEFFSMLAYASGGKPLLKLFLSRGDNI KW
YrhP Bsu (125) SNILNPKTVLVYVTIMPQFINLNGN INQQLII LASILTLLAVLWFLFLVYIIDYAKKW MKNSKFQKV
YfiK Eco (113) LQFVNVKIILYGVTALSTFVLPQT  QALSWVV  GVSVLLAMIGTFGNVCWALA GHLFQRLFRQYGRQ

Consensus      --L-NPK--LF--S----F------ ------- ------L-------------------R--------

YigJ Eco (187) IDGFAGALFAGFGIHLIISR
YigK Eco (183) LNKIFGSLFMLVGALLASARHA
YcaR Pae (195) ATSGVGALFVGFGVKLATATLS
Yg27 Syn (187) MNYLAGALMISVGVFLLISS
YahN Eco (200) ASRVIGAIIGVFALRLIYEGVTQR
YeaS Eco (192) GNSLIGLMFVGFAARLATLQS
Yd07 Hin (185) IDNMAGIVFLFFGCVLVYNGINEIIH
MlgA Sco (188) MNRIAGSLMICVGLWLALG
YrhP Bsu (192) FQKITGIILVGFGIKTGLS
YfiK Eco (178) LNIVLALLLVYCAVRIFY

Consensus      -----G-L---FG--L-----
```

Figure 7. Multiple alignment of ten members of the YahN family. Protein abbreviations are as provided in Table 2. The alignment was generated using the TREE program. The conventions of presentation are as for Figure 1.

Table 3. Protein Members of the CadD Family

Abbreviation	Name or description	Organism	Size (no. residues)	Database and accession no.
CadB Slu	Cadmium resistance protein	*Staphylococcus lugdunensis*	209	gbU74623
CadD Sau	Cadmium resistance protein	*Staphylococcus aureus*	209	gbU76550
QacF Bfi	Putative quaternary amine transporter	*Bacillus firmus*	197	gbZ17326
Orf Sau	Uncharacterized	*Staphylococcus aureus*	219	gbL10909

The situation with respect to the amino terminal part of LysE, representing about one quarter of the polypeptide chain and corresponding to hydrophobic regions #1 and #2, is more complex. In accordance with the experimental analysis resulting in assignment of the orientations of hydrophobic segments #3 to #6, putative loop 1:2 should be oriented towards the inside (Figure 6A). However, fusions of PhoA at positions 29 and 33 resulted in translocation of alkaline phosphatase to the periplasm. Thus, the amino terminal polypeptide of 29 amino acyl residues can replace the natural hydrophobic leader of alkaline phosphatase. Since a lysyl residue is located at position 30 in LysE, which is immediately followed by another positively charged residue (arginyl), two additional fusions were made at positions 37 and 40. This was done because positively charged residues can assist in the final orientation of helices in the membrane as demonstrated for the leader peptidase, Lep, of *E. coli* (v. Heijne, 1994). However, translocation of PhoA was still observed for the LysE-PhoA fusions which included these positively charged residues. Therefore, hydrophobic region #1 serves as a topological determinant for PhoA translocation, and the putative loop between hydrophobic regions #1 and #2 is directed towards the periplasm. A plausible model based on the fusion analyses is presented in Figure 6B.

The YahN Family

All of the proteins of the LysE family were screened using the BLAST program in an attempt to find additional homologues of the established proteins of the LysE family. One protein appeared with a poor score. This protein is YahN of *E. coli*. When YahN was BLASTed against the databases, it proved to be a member of a family of substantial size. The ten proteins identified are listed in Table 2. They all are in the same size range as members of the LysE family. Most of the proteins of the YahN family are from Gram-negative bacteria, five from *E. coli*. However, one of the YahN family members is from the Gram-positive bacterium *B. subtilis* (YrhP), and one is from the cyanobacterium *Synechocystis* PCC6803 (Yg27). All ten

```
                                                                    *  *
Orf Sau     (1)  MKKSPIPKIFGDDKMIATILTATAVYVATGIDYLVILILLFSQVK KG QVKHIWIG
CadB Slu    (1)   MRCNYDSDYCN          CCYLYIAQALDLLVILLMFFARAKTRK EYRDIYIG
CadD Sau    (1)   MRCIMIQTVVA          AAVLYIATAVDLLVILLIFFARAKTRK EYRDIYVG
QacF Bfi    (1)                        MLLFSQAKAQVKTSEGLKGNQTENKHISPKDIIIG

Consensus          -----------          ---LY-A---D-LVIL---F---K-------DI--G


                  ** *       **          *  *  ***** * ***
Orf Sau    (56)  QYIGTAIVIGASLLVAQGVVNLIPQQWVIGLLGLLPLYLGVKIWIKGEEDEDESSIL
CadB Slu   (47)  QYVGSVALIVISLFFA FVLNYVPEKWILGLLGLIPIYLGIKVAIYGDSDGEERAKK
CadD Sau   (47)  QYLGSIILILVSLPLA FVLNYVPEKWILGLLGLIPIYLGIKVAIYDDCEGEKRAKK
QacF Bfi   (36)  QYLGFTLLVVVSLLGT FGVMLIPEKWV GLLGLIPIYLGIMLFIKGEDEDENAILS

Consensus        QYLG---LI--SL--A-F-VN--PEKW--GLLGLIPIYLGIK--I-G----E-----


                  *   *                    *       * * **
Orf Sau   (113)  SLFSSGKFNQLFLTMIFIVLAS SADDFSIYIPYFTTLSMSEIFIVTIVFLIMVGVL
CadB Slu  (103)  ELNEKG LSKLVGTIAIVTIASCGADNIGLFVPYFVTLSVTNLLITLFVFLILIFFL
CadD Sau  (103)  ELNEKG LSKLVGTVAIVTISSCGADNIGLFVPYFVTLSVTNLLLTLFVFLILIFFL
QacF Bfi   (91)  SL NSGKYNSVFLSVAFITFAN GGDNIGIYVPFFSTLTMNQLAVTVITFFIMVAIW

Consensus        -L---G----L--T-A--T--S-GADNIG--VPYF-TLS---L--T--VFLI----L


                       **      *  *   ***      * **  *  ** *
Orf Sau   (169)  CYVSYRLASFDFISETIEKYERWIVPIVFIGLGIYILFENGTSNALISFL L
CadB Slu  (159)  VFAAQKLANIPEVGEIVEKFGRWIMAVIYIALGLFIIIENDTIQTILGFI F
CadD Sau  (159)  VFTAQKLANIPGIGEIVEKFSRWIMAIIYIALGLFIIIENDTIQTILGFI F
QacF Bfi  (146)  CFIGYRLAAFKHVSETLENYGRWIIPIVFIGLGIYIMVENETFSALLSLINY

Consensus        -F----LA------E--EK--RWI--I--I-LG--I--EN-T----L-FI -
```

Figure 8. Multiple alignment of the CadD family. Protein abbreviations are as described in Table 3, and the conventions of presentation are as indicated in Figure 1.

of the YahN family members were predicted to be related to the LysE family proteins based on PSI-BLAST results (Altschul *et al.*, 1997).

A multiple alignment of the ten members of the YahN family is shown in Figure 7. Only four residues are fully conserved, revealing a very significant degree of sequence divergence among these proteins. The average similarity analysis (not shown) revealed that hydrophobic regions #2, #3, #4 and #6 (see Figure 2B) are all well conserved. Hydrophobic region #4, which is less hydrophobic than hydrophobic regions #2, #3 and #6, is the most strongly conserved portion of YahN family proteins (compare Figure 7 with Figure 2B). The phylogenetic tree for the ten members of the YahN family analyzed revealed little clustering (Figure 3B). Based on the recognized phylogenetic relationships of the organisms, we suggest that YeaS Eco and YcaR Pae may be orthologues, but that YigJ Eco and Yd07 Hin are not.

The CadD Family

Results obtained with the PSI-BLAST program suggested that the LysE and YahN families are related to a third more distant family of proteins that we have called the CadD family (Table 3). Two of the four sequenced proteins that comprise this family are plasmid-encoded and probably function in cadmium resistance (Chaouni *et al.*, 1996). A third member of this family has tentatively been suggested to catalyze efflux of quarternary amines (see the database entry of this protein). The fourth protein has not been examined for function. All four of these proteins are derived from Gram-positive bacteria.

The multiple alignment of the four CadD family members is shown in Figure 8. Many fully conserved residues are found within this set of proteins, consistent with the conclusion that they do, in fact, comprise a distinct family within the LysE superfamily. The average hydrophobicity plot (Figure 2C) reveals a pattern strikingly similar to those noted above for the proteins of the LysE (Figure 2A) and YahN (Figure 2B) families. The phylogenetic tree for the CadD family (Figure 3C) revealed that the two orthologous cadmium resistance proteins cluster tightly together while the QacF Bfi and Orf Sau proteins are about equidistant from the CadB and CadD proteins. Proteins of the CadD family exhibit no sequence or motif similarity to other recognized heavy metal ion exporter families (Paulsen and Saier, 1997).

Discussion

LysE Topological Model

Sequence analyses of LysE superfamily proteins revealed six hydrophobic regions (Figure 2A-C). Our experimental fusion protein analyses of the *C. glutamicum* LysE protein showed that hydrophobic regions #3 to #6 each traverse the membrane once, placing loops 3:4 and 5:6 towards the cytoplasmic side, and loop 4:5 and the distal end of hydrophobic region #6 towards the periplasmic side of the membrane (Figure 6). Further experimental results showed that the expected arrangement of hydrophobic regions #1 and #2 of LysE in an alternating transmembrane array is not valid.

Instead, the experimental data obtained from LysE'PhoA fusions 29, 33, 37, 40, 59, 67 and 75 place these residues in LysE in the periplasm. Limitations of topological interpretation for reporter gene fusions near the N-terminal end of a protein have been reported. Thus, fusions can prevent translocation (Tate and Hendersen, 1993; Turk *et al.*, 1996) or even reverse orientation of a membrane spanning helix due to missing positive charges (v. Heijne, 1994). Importantly, in both cases formation of active PhoA is prevented. However, in the case of the LysE'PhoA fusions, alkaline phosphatase is active, showing that translocation occurred. Therefore, hydrophobic region #1 is believed to be a transmembrane spanning helix with its carboxyl end directed towards the periplasm. Based on these results we propose that hydrophobic region #2 is not membrane spanning. Instead it either might be located peripherally on the periplasmic side of the membrane or loop into the membrane. Further evidence for an unusual topology for the N-terminal part of LysE is the rather short putative helix #1, which probably does not extend beyond the membrane, and the fact that loop 2:3 is not appreciably amphipathic (not shown). Although further investigations on the structure at the N-terminal part of LysE are certainly necessary, nevertheless the 2-D structure is conclusive for about three-quarters of the protein. Recently obtained 3-D structures of membrane proteins, such as those of aquaporin (Walz *et al.*, 1997), and of K^+ channels (Nelson *et al.*, 1999) show that regions are present which do not simply form transmembrane spanning helices or loops and which are mechanistically of great significance.

Possible Functionally Important Residues

The identification of conserved residues in LysE together with a helical wheel analysis (not shown) allows us to speculate on residues which might be of functional significance. The best conserved region (LNPHVYL) is localized to the center of hydrophobic region #4 which is proposed to comprise a central part of the translocation channel. Neither the hydrophilic residues nor the fully conserved residues of hydrophobic region #4 are localized to one side of the helix. However, an important function for the conserved LNP motif can be postulated. The fully conserved asparagine within this motif is one helical turn away from a conserved threonine/serine residue, and these polar/semipolar residues could function together as part of the transmembrane active site. The location of the prolyl residue within this motif is in full accord with this view. Prolyl residues in membrane spanning helices of bacteriorhodopsin introduce kink angles of about 20° to position functionally important residues in the three dimensional structure (Grigorieff *et al.*, 1996). For example, Pro50 in helix B in bacteriorhodopsin positions an unpaired carbonyl oxygen of Thr46 which forms part of the channel (Deisenhofer and Michel, 1989). Importantly, the prolyl and threonyl residues in bacteriorhodopsin are separated by 4 residues as is the case for these residues in LysE. These residues, together with the aspartate in hydrophobic region #2, could play a role in lysine binding and/or in binding of a putative cotransported or countertransported cation.

The L-lysine exporter of *C. glutamicum* has been characterized as a secondary active transport system where lysine translocation is thought to be coupled to H^+ influx or OH^- efflux (Broer and Krämer, 1991). The relatively well conserved D/E and H residues in hydrophobic region #4 are localized near one another on one side of this putative helix. The two fully conserved aromatic residues

of hydrophobic region #5 are localized to the same side of this putative helix, with the tryptophanyl residue preceeded by and one helical turn away from the well conserved serine. Finally, hydrophobic region #6 exhibits one fully conserved M residue, and a well conserved D/N residue, two helical turns away. We therefore suggest that (a) the fully conserved D in hydrophobic region #2, (b) the N and T/S residues in hydrophobic region #4, (c) the fully conserved aromatic residues in hydrophobic region #5 and (iv) the D/N and M residues in hydrophobic region #6 represent reasonable candidates for residues that might line the transmembrane channel.

A Novel Superfamily of Solute Exporters?

The intensive sequence analyses reported in this communication serve to define a novel superfamily (the LysE superfamily) of distantly related homologues which comprise three distinct families, the LysE, YahN and CadD families. The evidence that these proteins belong to a single superfamily can be summarized as follows: (i) all of the proteins of these three families have essentially the same sizes ranging from 195 to 236 amino acyl residues; (ii) all of these protein families exhibit essentially the same average hydrophobicity plots, a fact that suggests that the constituent proteins exhibit very similar three-dimensional structures; (iii) functionally characterized members of the LysE superfamily catalyze solute efflux; (iv) proteins of the LysE and YahN families exhibit convincing degrees of sequence similarity; (v) the PSI-BLAST program grouped all of the proteins of the LysE, YahN and CadD families together in a single superfamily (data not shown); (vi) none of the proteins predicted to be included in the LysE superfamily could be found in eukaryotes.

A surprising fact emerges from the data summarized in Tables 1-3. Although only one of the LysE family homologues is from *E. coli*, and none of the CadD family members is from *E. coli*, five of the ten YahN proteins are from *E. coli*. Why this one bacterium would proliferate so many YahN paralogues poses an interesting question. The fact that the YahN members of *E. coli* have not been functionally investigated provides an intriguing argument in favour of a novel function for these proteins.

Note Added in Proof

After completion of this manuscript, the functions of two members of the YahN family were reported (Aleshin, V.V., Zakataeva, N.P., and Livshits, V.A. 1999. A new family of amino acid efflux proteins. Trends Biochem. Sci. 24: 133-135; Zakataeva, N.P., Aleshin, V.V. Tokmakova, I.L. Troshin, P.V., and Livshits, V.A. 1999. The novel transmembrane *Escherichia coli* proteins involved in amino acid efflux. FEBS Lett. 452: 228-232). The two proteins are YigK, renamed RhtB, a probable homoserine/homoserine lactone/ß-hydroxynorvaline exporter, and YigJ, renamed RhtC, a probable threonine exporter. We have therefore renamed the YahN family, the RhtB family. The three constituent families of the LysE superfamily are therefore (1) the LysE family (TC *#2.75)*, (2) the RhtB family (TC #2.76) and (3) the CadD family (TC #2.77) (see our web site http://www-biology.ucsd.edu/~msaier/transport/).

Experimental Procedure

Fusion Constructs

Fragments to be inserted into pPA4 which contains *phoA* devoid of its promoter and leader peptide were generated via PCR using pSPORT*lysE* as a template (Vrljic *et al.*, 1996). As a sense primer, 5'-CTGTCCTGCAGCTTCCATAGGTCACGATGGTG-3' was used which anneals at the 5'end of *lysE* and generates a *Pst*I restriction site. Twenty two antisense primers were selected along the *lysE* sequence, generating suitable restriction sites for *Hpa*I, *Pma*CI or *Eco*RV to enable the construction of inframe fusions. PCR amplified fragments were digested and cloned into pUC18 before cloning into pPA4. The following fusions were made (where numbering refers to the last amino acid specific to LysE): 29, 33, 37, 40, 59, 67, 75, 99, 111, 112, 113, 114, 115, 120, 130, 243, 154, 164, 171, 183, 211, and 236. The correctness of the fusions was confirmed by sequencing (except fusions 111-130). After construction of the plasmids in *E. coli* DH5αmcr, plasmids were introduced into *E. coli* CC118 which lacks endogenous alkaline phosphatase activity (Manoil and Beckwith, 1986). To make *lacZ* fusions the pGP vectors of Haardt and Bremer (1996) were used, as well as the appropriate primers to enable fusions at position 40, 67, 111, 171, and 211. The phenotype of the fusions 111 and 211 obtained was estimated in *E. coli* CC118 on LB plates containing 5-bromo-4-chloro-3-indolyl-ß-D-galactoside (40 µg/ml).

Enzyme Activity Determinations and Immunodetection of Fusion Proteins

Single colonies of recombinant CC118 strains were grown overnight at 37°C in Luria-Bertani medium supplemented with 50 µg/ml carbenicillin. These cultures were used to inoculate new cultures which were grown for 3 hours to an OD of 1. Cells were permeabilized and used for phosphatase activity determinations based on an extinction coefficient of $\varepsilon_{405\ nm}$=0.0188/M/cm for the chromogenic product 5-bromo-4-chloro-indolyl phosphate. For Western-blot analyses, cultures were grown identically, but after harvesting, they were washed with 10 mM TrisHCl, 10 mM $MgSO_4$ (pH 8) and disrupted by sonification in the same buffer containing 40 µl/ml protease inhibitor Complete™ (Roche). Each sample (150 µg) was separated by SDS/PAGE and blotted onto nitrocellulose membranes. The PhoA protein was probed with monoclonal antibody DC133A (5'Prime Inc.; USA) and visualized with anti-IgG-coupled alkaline phosphatase using 5-bromo-4-chloro-indolyl phosphate/nitro blue tetrazolium as substrate.

Computer methods

The BLAST and PSI-BLAST data base search methods have been described (Altschul *et al.*, 1997). The GAP program of Devereux *et al.* (1984) was used to establish homology and determine relative degrees of sequence similarity. Multiple alignments, as well as phylogenetic trees were generated using the PREALIGN and TREE programs (Feng and Doolittle, 1990). Average hydropathy, similarity and amphipathicity plots based on the TREE-generated multiple alignments were produced as described (Kyte and Doolittle, 1982; Le *et al.*, 1999) using a sliding window of 21 residues. Data base sequence entries were corrected as indicated in the footnotes to Tables 1 and 2.

Acknowledgements

We thank U. Reuter and M. Simonaitis for assistance in the preparation of this manuscript and K. Krumbach for PhoA activity determinations. Work in the laboratory of M. Saier Jr. was supported by USPHS grant 5RO1 AI21702 from the National Institutes of Allergy and Infectious Diseases, grant 9RO1 GM55434 from the National Institute of General Medical Sciences and by the M.H. Saier, Sr. Memorial Research Fund. Work in the laboratory of H. Sahm was supported by grants from DEGUSSA AG, Frankfurt, and Bundesland Nordrhein-Westfalen, Germany.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25: 3389-3402.

Bröer, S., and Krämer, R. 1991. Lysine excretion by *Corynebacterium glutamicum*: Energetics and mechanism of the transport system. Eur. J. Biochem. 202: 137-143.

Chaouni, L.B., Etienne, J., Greenland, T., and Vandenesch, F. 1996. Nucleic acid sequence and affilation of pLUG10, a novel cadmium resistance plasmid from *Staphylococcus lugdunensis*. Plasmid. 36: 1-8.

Deisenhofer, J., and Michel, H. 1989. Nobel lecture: the photosynthetic reaction centre from the purple bacterium *Rhodopseudomonas viridus*. EMBO J. 8: 2149-2169.

Devereux, J., Haeberli, P., and Smithies, O. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12: 387-395.

Doyle, D.A., Cabral, J.M., Pfuetzner, R.A., Kuo, A., Gulbis, J.M., Cohen, S.L., Chait, B.T., and MacKinnon, R.M. 1998. The structure of the potassium channel: Molecular basis of K+ conduction and selectivity. Science. 280: 69-77.

Eggeling, L. 1994. Biology of L-lysine overproduction with *Corynebacterium glutamicum*. Amino Acids. 6: 261-272.

Eng , B.H., Guerinot, M.L., Elde, D., and Saier, M.H., Jr. 1998. Sequence analyses and phylogenetic characterization of the ZIP family of metal ion transport proteins. J. Membr. Biol. 166: 1-7.

Feng, D.F., and Doolittle, R.F. 1990. Progressive alignment and phylogenetic tree construction of protein sequences. Methods Enzymol. 183: 375-387.

Grigorieff, N. Ceska, T.A., Downing, K.H., Baldwin, J.M., and Henderson R. 1996. Electron-crystallographic refinement of the structure of bacteriorhodopsin. J. Mol. Biol. 259: 393-421.

Haardt, M., and Bremer, E. 1996. Use of *phoA* and *lacZ* fusions to study the membrane topology of ProW, a component of the osmoregulated ProU transport system of *Escherichia coli*. J. Bacteriol. 178: 5370-5381.

Kyte, J. and Doolittle, R.F. 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157: 105-132.

Le, T., Tseng, T., and Saier, M.H., Jr. 1999. Flexible programs for the estimation of average amphipathicity of multiply aligned homologous proteins: application to integral membrane transport poteins. Mol. Mem. Biol. 16: 173-179.

Leuchtenberger, W. 1996. Amino acids - Technical production and use. *In*: Products of primary metabolism (eds Rehm, H.J. and G. Reed) Biotechnol. 6: 455-502.

Manoil, C., and Beckwith, J. 1986. A genetic approach to analyzing membrane protein topology. Science. 233: 1403-1408.

Nakayama, K. 1985. L-Lysine. In: Comprehensive Biotechnology vol. 3. M. Moo-Young, ed. Oxford University Press, Oxford. p. 607-620.

Nelson, R.D., Kuan, G., Saier, M.H. Jr., and Montal, M. 1999. Modular assembly of voltage-gated channel proteins: a sequence analysis and phylogenetic study. J. Mol. Microbiol. Biotechnol. 1: 281-287.

Paulsen I.T., Skurray, R.A., Tam, R., Saier M.H. Jr., Turner, R.J., Weiner, J.H., Goldberg, E.B., and Grinius, L.L. 1996. The SMR family: a novel family of multidrug efflux proteins involved with the efflux of lipophilic drugs. Mol. Microbiol. 19: 1167-1175.

Paulsen, I.T., and Saier Jr. M.H. 1997. A novel family of ubiquitous heavy metal ion transporter proteins. J. Membrane Biol. 156: 99-103.

Saier, M.H. Jr. 1994. Computer-aided analyses of transport protein sequences: Gleaning evidence concerning function, structure, biogenesis, and evolution. Microbiol. Rev. 58: 71-93.

Saier, M.H. Jr., Tam, R., Reizer, A. and Reizer, J. 1994. Two novel families of bacterial membrane proteins concerned with nodulation, cell division and transport. Mol. Microbiol. 11: 841-847.

Saier, M.H. Jr. 1998. Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea, and eukarya. Adv. Microbiol. Physiol. 40: 81-136.

Sarsero, J., and Pittard, A.J. 1995. Membrane topology analysis of Escherichia coli K-12 Mtr permease by alkaline phosphatase and B-glactosidase fusions. J. Bacteriol. 177: 297-306.

Tate, C.G., and Henderson, P.J. 1993. Membrane topology of the L-rhamnose-H+ transport protein (RhaT) from enterobacteria. J. Biol. Chem. 268: 26850-26857.

Turk, E., Kerner, C.J., Lostao, M.P., and Wright, E.M. 1996. Membrane topology of the human Na+/glucose cotransporter SGLT1. J. Biol. Chem. 271: 1925-1934.

v. Heijne, G. 1994. Membrane proteins: From sequence to structure. Ann. Rev. Biophys. Biomol. Struct. 23: 167-192.

Vrljic, M., Eggeling, L., and Sahm, H. 1995. Unbalance of L-lysine flux in *Corynebacterium glutamicum* and its use for the isolation of excretion defective mutants. J. Bacteriol. 177: 4021-4027.

Vrljic, M., Eggeling, L., and Sahm, H. 1996. A new type of transporter with a new type of cellular function: L-lysine export from *Corynebacterium glutamicum*. Mol. Microbiol. 22: 815-826.

Walz, T., Hirai, T., Murata, K., Heyman, J.B., Mitsuoka, K., Fujiyoshi, Y., Smith, B.L., Agre, P., and Engel, A. 1997. The three-dimensional structure of aquaporin-1. Nature. 387: 624-627.